Aim: To detect pneumonia from X-ray images of lungs

Background:

Corona virus may spread through respiratory airways reaching the lungs, where it may lead to the development of severe viral pneumonia, identified as the final death cause in most lethal cases. Lung monitoring for pneumonia has gained greater importance during Corona outbreak.

Pneumonia is an inflammatory condition of the lungs, primarily affecting the small air sacs known as alveoli. Symptoms typically include some combination of productive or dry cough, chest pain, fever and difficulty breathing. Pneumonia is usually caused by infection with viruses or bacteria and less commonly by other microorganisms, certain medications or conditions such as autoimmune diseases.

Diagnosis is often based on symptoms and physical examination. Chest X-ray, blood tests, and culture of the sputum may help confirm the diagnosis. Along with physical examination, imaging diagnosis plays a central role in the detection of pneumonia. Chest radiographs are frequently used in diagnosis procedures and represent a fast, cost-effective alternative to map the nature, features and extension of lung inflammations. X-ray radiograph image opacity areas are commonly correlated to pneumonia affected regions.

The identification of opacities in chest x-ray images is sometimes challenging. Artificial intelligence and computer vision techniques have been extensively used over the past few years in pneumonia diagnosis. In this project, we will go through how to detect pneumonia from X-ray images. This helps in early detection of pneumonia in a faster and efficient manner so that Corona patients having pneumonia can be given a top priority.

The dataset:

Link: https://www.kaggle.com/pcbreviglieri/pneumonia-xray-images

Pre-processing:

The input shape is a 400x400 pixels image with 1 channel, considering grayscale images.

Keras 'ImageDataGenerator' is used to generate augmented training and validation image sets. Two instances of this object are created:

'train_gen' will be utilized to generate the training set. rescaling is performed through the 'rescale' parameter and augmentation is achieved with some image shearing, zooming and vertical flips.

'val_gen' will be utilized to generate the validation set. Only rescaling is performed in this case, as we just need to test the model for accuracy.

Code Description:

Kernel : 3x3x1

Activation function : 'ReLU' (Rectifier Linear Unit)

Number of epochs : 50

Model :

Three pairs of alternate convolution and pooling layers:

The first and second pair produce 32 feature maps processed by subsequent 2x2 pooling, while the third produces 64 feature maps also processed by a subsequent 2x2 pooling.

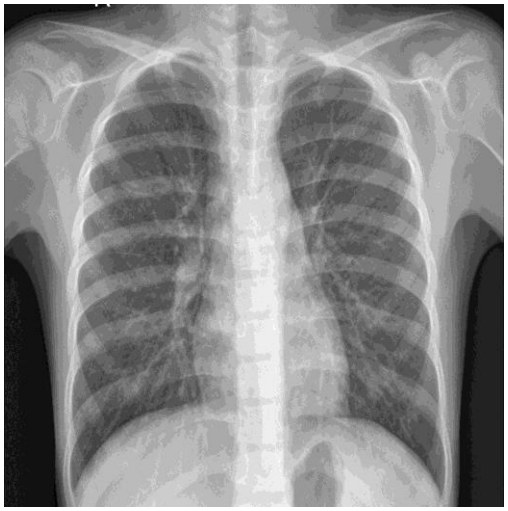
Fully connected Dense layers with 64 and 1 neuron, respectively, perform the final binary classification.

Predicting pneumonia from test set:

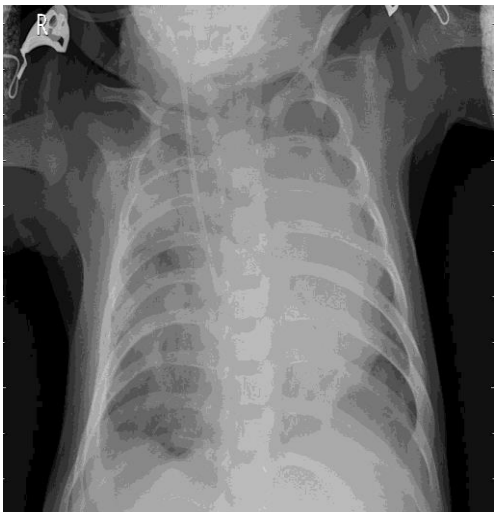
The model will return the probability of pneumonia for each image. Higher the output, the higher the probability of a pneumonia occurrence. For the purpose of classification, use a threshold of 0.5 to get corresponding 'normal' and 'pneumonia' diagnosis.

Results:

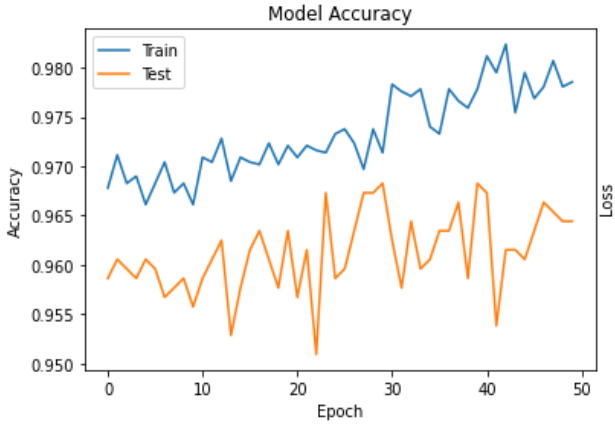
X-ray scans:

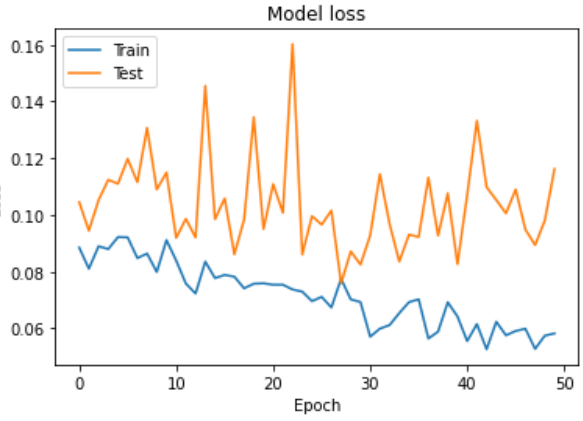


Normal



Infected





Confusion Matrix:

	Predicted Normal	Predicted Pneumonia
Actual Normal	186	48
Actual Pneumonia	9	381

Accuracy: 90.86538461538461

The case of actually being normal but predicted as Pneumonia is quite higher than expected. One reason could be the model detected less opacity in 'normal' scans and classified them as infected. More experiments could provide optimum batch size, dimensions and epochs.